
Loney, George. "The University of Guelph Library's SearchMe
Public-Access Catalogue." The Public-Access Computer Systems
Review 1, No. 3 (1990): 30-43.

1.0 Introduction

The University of Guelph is a medium-sized university located in southwestern Ontario about 100 kilometers from Toronto. The library has been automating its various systems since the mid-1960s, starting with electronic data collection devices for a batch-oriented circulation system. The systems that followed included a batch cataloguing system called Scope, the CODOC system, and the Geac online circulation system (co-developed with Geac). The Geac circulation system was expanded to include online public access, acquisitions, and cataloguing, all running on the Geac mini-computers.

In 1987, the University of Guelph Library began a pilot project to determine the viability of individual CD-ROM workstations as a replacement for its centralized online catalogue. This storage medium for the nearly 900,000 record bibliographic database was chosen because it offered an extremely cost-effective method of distributing the 500-megabyte database to what is projected to be a network of over 100 workstations.

The original version of the search software and database was the product of a commercial vendor. The pilot project determined that while CD-ROM was an acceptable medium for storing and retrieving the data, the software used during the pilot project was not desirable for the long term, and the inability to change the database would require frequent and costly remasterings.

As a result, a database design was developed and tested that would allow the library to write its own search software, prepare its own database, deal directly with the CD-ROM manufacturers at a greatly reduced cost, and add changes to the CD-ROM data. This software project was started in May 1988, and the new system was installed in October 1988 on 25 workstations throughout the library. Since then, the system has completely replaced the old, centralized online public access system and is running on 85 workstations in the two library branches and on a few additional workstations in some academic departments.

This article will examine some of the issues surrounding the development of the SearchMe software relating to the user interface and implications of the use of the CD-ROM as the major storage medium.

2.0 User Survey

Prior to the development of SearchMe, a survey of library users was conducted by the systems staff with the help of the reader service staff. Patrons were approached while they were using one of the publicly available search tools: the card catalogue (we still had one at the time), the online public access system, or the circulation inquiry system. Questions were asked to determine what information patrons had when they started a search, what it was they were looking for, and how well or poorly the current search tools satisfied them.

A number of conclusions were inescapable:

1. Patrons learn how to use the library systems through different means, but self-teaching is the most usual method.
2. While patrons are concerned if system response time is slow, they become very frustrated when response time is inconsistent, e.g., when use is heavy.
3. Patrons migrate very easily from the card catalogue to computer-supported search tools. The only difficulty with the search tools is that many terminals or workstations are needed to prevent line ups.
4. Patrons using the automated search tools perceived that they had found most or all the available information. We were never able to establish how they knew that they had found "all" the information, but it was indicative of their perception that they were being adequately helped.

As a result of this and other knowledge sources, we developed a design goal where the new system would attempt to:

1. Provide highly consistent response times no matter how high the user load or how many terminals were in the overall system.
2. Provide high functionality first and high speed second.
3. Be very consistent in its user interface and as intuitive as possible in its control functions.

+ Page 32 +

4. Provide context-sensitive help at all stages of system use.
5. Allow the novice user to become familiar with the search system with minimum formal instruction and permit the more experienced user to perform more complex searches.
6. Be very accurate in its information delivery and highly tolerant of user input error.

We believe that SearchMe is very successful at meeting these goals.

3.0 Consistent Response Time

SearchMe operates in a functionally distributed environment. Each workstation at the University of Guelph Library consists of a PC/XT clone with a 10 or 12.5 MHz 8088 chip, 640 KB of memory, an Ethernet card, one floppy drive, a 40 MB hard drive, an internal CD-ROM player, a monochrome monitor, and a rugged keyboard. There is a custom, lockable, front panel that covers the hard disk and CD-ROM player openings as well as blocks off the reset and turbo switches.

The minimum hardware requirement for SearchMe is an XT with 640 KB of memory and a single floppy drive. The software will take advantage of colour monitors if they are present, and it will alter certain display characteristics for colour monitors.

To reduce dependence on server or other response-time bottlenecks in the LAN, we make little use of the local area network. Changes to the catalogue database are transported automatically to the workstations during the night via the LAN. The workstation detects and transports software changes on start up, and requests for circulation information about patron or bibliographic records are handled by the LAN. If the LAN or the server is inoperative, the software recognizes this condition, and the affected functions are simply declared unavailable.

+ Page 33 +

The key to consistent response times is the fact that each workstation contains the entire library catalogue database and its indexes on one resident CD-ROM. There is a limit of about 600-650 MB of data that can be put on a CD-ROM. We have our entire collection of about 900,000 bibliographic records on one CD-ROM disc, and we believe we can expand our database to about 1.2 million records without adding a second CD-ROM. If this possibility occurred (rather remote given current acquisitions budgets), we have several options: the text data could be compressed to reduce the amount of space required, machines could be twinned to share CD-ROM players, or machines could be clustered around a data server.

Another advantage of a self-contained system is that functions that could previously be provided to users only with large (and expensive) centralized processors, are now possible with a microcomputer-based CD-ROM system since the computing resource is not shared by anyone else. Boolean searches on large collections of data can be provided with no penalty to the rest of the system.

4.0 Functionality

As many functions as possible were considered in the design of SearchMe. Functions were rejected only if they were too complicated or were useful to only a very small group of users. As a result, the types of searches available on the system are: (1) full title search, (2) full author search, (3) full call number search, and (4) subject search. Subject search allows patrons to access data using: (1) titles; (2) corporate and personal authors; (3) call numbers; (4) Library of Congress Subject Headings; (5) material type names in the detailed holdings statements; (6) location names from the detailed holdings statements; (7) collection names; (8) any word from either the title, author, or subject heading fields; and (9) any word from most places in the record. These access points can be combined using the Boolean operators "AND," "OR," or "NOT."

The full title, author, and call number searches allow a simple, single phrase search that our survey showed most people use to find much of the material they want.

A further feature allows users to shelf browse forward and backward from any record they find. This capability closely corresponds to browsing the actual shelf because the database is organized in shelf sequence.

+ Page 34 +

Users may also display search results on the screen, request a printout of the results, or save them as an ASCII file on a floppy diskette. Users may customize the output as they wish, and they may print, display, or save any result record.

In addition, the system can link directly to current circulation status information so that users may request display of their own current biographical information, including items on loan, overdue fines, outstanding holds, and available holds. The system allows patrons to place holds on items and will automatically transfer them to the Circulation System.

5.0 Consistent User Interface

In keeping with current user interface practice, a highly consistent interface has been implemented.

The top of the screen is used to display messages about the current status of the search in progress; the middle of the screen is used to display index lists, search strategies, search results, and detailed help; and the bottom of the screen contains short directions to the user and error messages.

No key is used for two different kinds of function command, and a special set of coloured key caps has been installed with customized legends (e.g., find by title, find by author, help, next record, and previous record). Assigning custom key caps has

freed the screen for anecdotal directions (e.g., "Press one of the blue keys to start your search") instead of messages that are concerned solely with keyboard use. The largest key on the keyboard, coloured bright red, is the help key. When a user presses this key, a window pops up containing a description of the current screen. The amount of text that can be displayed in this window is unrestricted.

+ Page 35 +

6.0 Learning the System

As our survey found, there are many ways that people learn to use the system. At the beginning of each semester, the library provides orientation classes that cover all the facilities available to our patrons. However, many users simply sit down at a terminal and start to use the system. As a result, we have made specific provisions for this type of approach. Use of the system itself is largely intuitive; the commands are printed right on the key caps. Using these and the screen prompts, many patrons can start doing simple searches without any previous instruction. Located at the various workstations are one-page instruction sheets that explain the purpose of the function keys and the contents of the index access points. Also available are scripts that lead the user through a sample search.

7.0 Searching

To perform a simple search, the user presses the Find By Title key, the Find By Author key, or the Find By Call Number key (see Figure 1).

Figure 1. Main Screen

The University of Guelph Library
Catalogue Access System

Press one of the blue keys to do a simple search. Press the subject search key to perform combined term searches or to access indexes other than the title, author, or call number.

+ Page 36 +

The system then prompts the user to enter the title (see Figure 2), author, or call number and press Enter.

Figure 2. User Is Prompted to Enter a Title

The University of Guelph Library
Catalogue Access System

Enter Title:

Type in the text that you wish to find and press the enter key.
The system will search for the closest match to the text that
you have entered.

When the Enter key is pressed, the system uses the search term
entered to place the user as close as possible to the desired
index entry (see Figure 3). Users can press the cursor control
keys around the list of index entries until they have located the
correct title, author, or call number. Then they can press the
Display Result, Save Result, or Print Result keys to view, dump
to diskette, or print the records.

Figure 3. List of Titles; Second Line Highlighted

The University of Guelph Library
Catalogue Access System

Enter Title: the large scale structure of space

+-----+
+-Index List-----+
Large-Scale Sharing of Computer Resources 1
Large-Scale Structure of Space-Time 1
Large-Scale Structure of the Universe 3
Large-Scale Structures in the Universe 1
Large-Scale Superimposed Folds in Precambrian Rocks of... 1
Large-Scale Systems Modelling

+-----+
Press the up, down, PgUp, PgDn keys to manipulate the index
list. Display a highlighted record with the display result
key. Press help for more information.

+ Page 37 +

While displaying records (see Figure 4), users can press the Page
Up and Page Down keys to view multi-screen records. The Next
Record and Previous Record keys are used to display other records
in the result, and the Browse Forward and Browse Backward keys

allow users to shelf browse around a specific result record. The red Undo/Esc key moves the user back one step at a time, and the Start Again key cancels everything and returns the screen to the beginning (see Figure 1). Any of the Find By keys also stop everything and start a new search.

Figure 4. Selected Record

```
The University of Guelph Library
Catalogue Access System                      record 1 of 1
+-Bibliographic Window-----+
  Call Number      QC 173.59.S65 H38
  Title            The Large Scale Structure of Space-Time
  Author           Hawking, S. W.
  Edition          Cambridge (Eng.) University Press, 1973
  Contents         Bibliography: p.373-380
  Series Title     Cambridge Monographs on Mathematical Physics
  Detailed Holdings:
  Cpy Location     Mat'l Type      Call Number
  1  Science      Book            QA 173.59.S65 H38
+------+
Press cursor key to see more of the record.  Press next or
previous record to look at other records in the set.  Press a
browse key to browse forward or backward from this record.
-----
```

The Subject Search key initiates complex searches (see Figure 5). The user is asked to choose an index from a list of All Keywords, Title Keywords, Author Keywords, Subject Heading Keywords, Titles, Authors, Subject Headings, Call Numbers, Material Types, Locations, and Collection Names. The system then prompts the user to enter the appropriate text and press the Enter key. After the search is conducted, the user is shown a list of index entries with the closest entry highlighted. The user selects a specific entry by moving the highlight around with the Up, Down, Page Up, and Page Down keys and then presses the Enter key.

+ Page 38 +

Figure 5. Select Initial or New Access Point

```
The University of Guelph Library
Catalogue Access System
+-Select Access Point-----+
  All Keywords
  Title Keywords
  Author Keywords
  LCSH Keywords
  Material Type
  Location
  Collection Name
  Title
```


distributed, static database applications. At first glance, it would seem that a library catalogue is relatively static; after all only about five per cent of the database changes in one year. However, that five per cent represents over 40,000 records for a medium-sized university collection--a large number of changes by any measure.

In our case, SearchMe was replacing a true online system. As changes were made to the database, they were immediately available to library patrons at the public terminals. The new system would have to be able to be updated on a regular, timely basis. SearchMe meets this objective in the design of its index system and its hardware configuration.

+ Page 40 +

Each workstation is connected to an Ethernet LAN. Periodically, when the workstation is otherwise inactive, it checks the central data server to see if there are any changes to the database. If so, the changes are copied into the workstation's hard disk. These changes are logically merged into the original CD-ROM resident database so that the library patron never actually knows whether the information is being delivered from the database changes or the original database.

9.0 Error Tolerance

There are two aspects to the system's ability to be tolerant of user errors: (1) how does it deal with incorrect control function commands, and (2) how does it react when search text is misspelled?

In the first case, the system generates error messages that attempt to inform users that they have made an error and why.

In the second case, the data retrieval software converts upper-case characters to lower case in both the entered text and the indexed text. Any punctuation (except for the call number) is changed to a space, and multiple occurrences of spaces are compressed to one space. In the title index, certain leading words (i.e., "the," "le," "la," and "les") are dropped unless they are the only word that was entered. Quite often, misspelled words will still result in the correct index entry display since the index mechanism attempts to find the entry that is "close to" the search term.

10.0 Technical Details

Workstation software is written in the C language. We currently use Borland's Turbo C version 2.0. The screen management, text manipulation, and indexes were all written by our staff. The database generation software is also written in C and runs in the Unix environment. It, too, was written entirely by our staff.

We currently send the prepared data to Discovery Systems in Columbus, Ohio, to have the CD-ROM discs made.

+ Page 41 +

The indexing scheme, also designed by our staff, is very efficient in its use of space and provides excellent response times. Our current bibliographic database uses 331 MB of space. The total space used by all the indexes is 211 MB. Data indexed includes (1) 1,424,000 titles; (2) 602,500 authors; (3) 286,000 subject headings; (4) 829,000 call numbers; (5) 651,000 keywords; (6) 71,000 ISBNs and ISSNs; (7) 206,000 L.C. Card Numbers; (8) location names; (9) material types; and (10) collection names. We do not use character compression, although if space became a problem, we could.

The CD-ROM disc is a very good device for serial access. It transfers data at much the same rate as a good hard disk; however, it is a slow random access device. For instance, the average seek time of a hard disk is about 30 ms whereas the CD-ROM needs between 270 and 340 ms. For this reason, the indexing scheme is optimized for the peculiarities of the CD-ROM medium (fewer than two disc seeks are required to go from search term entry to the closest occurrence of the term). Another two seeks are required to access the complete bibliographic record and display it on the screen. The system also attempts to predict user behaviour and pre-read data in order to speed the process even more. When run with a hard disk for storage, the software works very well and has extremely good response time.

Because of the ability to update data on the CD-ROM, we normally create a new CD-ROM version only every eight months or so. This process costs us about \$2,000 US for 300 copies of the disc.

Every possible part of the SearchMe software was put into parameters. The parameters are pre-loaded and optimized so that SearchMe does not have to interpret the data. The parameters are loaded by a programme that runs under MS-DOS and checks that they are accurate and viable. Some of the features that are controlled by parameter are:

- o Size and location of the data display windows, plus the kind of outline and title of the window (if any).
- o Colour of the window outline, background, and text, and the colour and other attributes (e.g., flash and reverse video) of highlighted text. The programme alters these values if the system is using a monochrome monitor.
- o Prompts, error messages, field names, and help messages.
- o Format of the bibliographic record display.

+ Page 42 +

- o Whether or not commands will be entered using the special keyboard or pull-down menus.
- o The content of the pull-down menus.

SearchMe supports multiple databases. The databases can be stored on CD-ROM, internal hard disk, or centralized (or distributed) data server. If there are multiple databases, users are given a menu of available databases and asked which one they wish to access. Each database uses its own parameter file so it is possible to configure each one quite differently from the others. Using multiple parameter files (which contain the prompts and other instructional text), it is possible to support multilingual applications by creating a parameter file for each language, where they all reference the same bibliographic data.

11.0 The Future

SearchMe is only the first phase of the complete rewrite of our online library system.

In December 1989, the cataloguing system was installed using the same type of architecture--distributed microcomputers accessing the main catalogue from a centralized server. With this approach, a highly sophisticated set of tools is available to the cataloguer, such as full-screen editing, interactive error detection, online coding manual, and online syntax checking. As with the SearchMe catalogue access, the system is highly reliable because it is not necessary for the central server to be available for work to continue.

We are about to add a binding module to the system. The basis of our authority control system is already included in the system, and this will be fully implemented later this year. Work is just starting on the development of our new circulation system after which we will add acquisitions and serials control.

We have also experimented with a low-cost optical scanner that we will use to scan and translate contents pages of incoming journals. From this, a SearchMe database of our journals, indexed by title, author, and keyword, will be maintained.

+ Page 43 +

12.0 Summary

The advent of high-capacity, inexpensive, personal storage devices such as CD-ROM has made the development of practical, large database workstations possible. The movement away from a centralized super-mini or mainframe computer to functionally distributed microprocessor workstations has allowed the University of Guelph Library to provide a highly functional, cost-effective, flexible catalogue access system. Ultimately, it will offer us the ability to move much more quickly to take

advantage of technological changes that benefit our user community.

About the Author

George Loney
Staff Analyst
University of Guelph Library
Guelph, Ontario N1G 2W1
Canada
BITNET: GLONWY@COSY.UOGUELPH.CA

The Public-Access Computer Systems Review is an electronic journal. It is sent free of charge to participants of the Public-Access Computer Systems Forum (PACS-L), a computer conference on BITNET. To join PACS-L, send an electronic mail message to LISTSERV@UHUPVM1 that says: SUBSCRIBE PACS-L First Name Last Name.

This article is Copyright (C) 1990 by George Loney. All Rights Reserved.

The Public-Access Computer Systems Review is Copyright (C) 1990 by the University Libraries, University of Houston. All Rights Reserved.

Copying is permitted for noncommercial use by computer conferences, individual scholars, and libraries. Libraries are authorized to add the journal to their collection, in electronic or printed form, at no charge. This message must appear on all copied material. All commercial use requires permission.
